ECE 537 Exam 2

UNIVERSITY OF ILLINOIS
Department of Electrical and Computer Engineering

October 24, 2022

- This is a closed-book exam.
- You are allowed to bring one 8.5x11 sheet of handwritten notes (front and back).
- No calculators are allowed. Please do not simplify explicit numerical expressions.
- There are 100 points in the exam. Points for each problem are specified by the problem number.

Name: ______________________________

NetID: ______________________________

Possibly Useful Charts and Formulas

Dynamic Time Warping

$$A_{i,k} = \max\left(A_{i-1,k}, A_{i,k-1}, a_{i,k} + A_{i-1,k-1}\right)$$

Linear Prediction

$$s[n] = Ge[n] + \sum_{m=1}^{p} a_m s[n-m] = h[n] * x[n]$$

$$H(z) = \frac{G}{1 - \sum_{m=1}^{N} a_m z^{-m}} = \frac{G}{\prod_{k=1}^{N}(1 - p_k z^{-1})}$$

$$\mathcal{E} = \sum_{n=0}^{N-1} e^2[n] = \sum_{n=0}^{N-1}\left(s[n] - \sum_{m=1}^{p} a_m s[n-m]\right)^2$$

$$0 = \sum_{n=0}^{N-1}\left(s[n] - \sum_{m=1}^{p} a_m s[n-m]\right) s[n-k], \quad 1 \le k \le p$$

$$\vec{c} = \Phi \vec{a}$$

Hidden Markov Models

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} b_j(\vec{o}_t), \quad 1 \le j \le N,\ 2 \le t \le T$$

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(\vec{o}_{t+1}) \beta_{t+1}(j), \quad 1 \le i \le N,\ 1 \le t \le T-1$$

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{k=1}^{N} \alpha_t(k)\beta_t(k)}$$

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(\vec{o}_{t+1}) \beta_{t+1}(j)}{\sum_{k=1}^{N}\sum_{\ell=1}^{N} \alpha_t(k) a_{k\ell} b_\ell(\vec{o}_{t+1}) \beta_{t+1}(\ell)}$$

$$\tilde{\alpha}_t(j) = \sum_{i=1}^{N} \hat{\alpha}_{t-1}(i) a_{ij} b_j(\vec{x}_t)$$

$$c_t = \sum_{j=1}^{N} \tilde{\alpha}_t(j)$$

$$\hat{\alpha}_t(j) = \frac{1}{g_t} \tilde{\alpha}_t(j)$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{j=1}^{N}\sum_{t=1}^{T-1} \xi_t(i,j)}$$

$$\bar{U}_i = \frac{\sum_{t=1}^{T} \gamma_t(i)(\vec{o}_t - \vec{\mu}_i)(\vec{x}_t - \vec{\mu}_i)^T}{\sum_{t=1}^{T} \gamma_t(i)}$$

$$\bar{\mu}_i = \frac{\sum_{t=1}^{T} \gamma_t(i) \vec{o}_t}{\sum_{t=1}^{T} \gamma_t(i)}$$

1. (20 points) Consider using dynamic time warping DTW to align two utterances, each of which has only two (2) frames. Let $a_{i,j}$ be the similarity between the i^{th} frame of utterance 1 and the j^{th} frame of utterance 2. DTW finds an optimal sequence of alignments, $I = [i(1) = 0, \ldots, i(T) = 2]$, $J = [j(1) = 0, \ldots, j(T) = 2]$, according to

$$I^*, J^*, T^* = \arg\max \sum_{t=1}^{T} S(i(t-1), i(t), j(t-1), j(t)),$$

where

$$S(i(t-1), i(t), j(t-1), j(t)) = \begin{cases} 0 & i(t) = i(t-1), j(t) = j(t-1) + 1 \\ 0 & i(t) = i(t-1) + 1, j(t) = j(t-1) \\ a_{i(t),j(t)} & i(t) = i(t-1) + 1, j(t) = j(t-1) + 1 \\ -\infty & \text{otherwise} \end{cases}$$

Under what circumstances would this method prefer to align frame 2 of utterance 1 with frame 1 of utterance 2?

Solution: This method will align frame 2 of utterance 1 with frame 1 of utterance 2 only if there is a path that enters $a_{1,2}$ diagonally that has a higher score than either of the other two possible paths. Since there are only two frames in total, the path that enters $a_{1,2}$ diagonally has no other diagonal steps, so this path will be chosen only if

$$a_{1,2} > a_{1,1} + a_{2,2}$$

and

$$a_{1,2} > a_{2,1}$$

2. (20 points) Suppose we want to predict the n^{th} sample of a speech waveform from two other samples, $s[n-P]$ and $s[n-Q]$, thus

$$e[n] = s[n] - a_P s[n-P] - a_Q s[n-Q]$$

and we wish to find the values of a_P and a_Q that minimize

$$\mathcal{E} = \sum_{n=0}^{N-1} (e[n])^2$$

What values of a_P and a_Q minimize $\mathcal{E}$? If you write your answer in terms of any other vectors, matrices, or covariance functions, be sure to define them.

Solution: By the principle of orthogonality, $\mathcal{E}$ is minimized if and only if

$$\sum_{n=0}^{N-1} e[n]s[n-P] = 0$$

$$\sum_{n=0}^{N-1} e[n]s[n-Q] = 0$$

Substituting in the definition of $e[n]$, we find that $\mathcal{E}$ is minimized by

$$\begin{bmatrix} a_P \\ a_Q \end{bmatrix} = \Phi^{-1}\vec{c},$$

where

$$\Phi = \begin{bmatrix} \phi(P,P) & \phi(P,Q) \\ \phi(Q,P) & \phi(Q,Q) \end{bmatrix}, \quad \vec{c} = \begin{bmatrix} \phi(0,P) \\ \phi(0,Q) \end{bmatrix},$$

and

$$\phi(m,k) = \sum_{n=0}^{N-1} d[n-m]d[n-k]$$

3. (20 points) Consider an LPC-based speech synthesizer with no pitch prediction; thus the speech signal $s_k[n]$ is generated from an excitation signal $e_k[n]$ using only

$$s_k[n] = e_k[n] + \sum_{m=1}^{p} a_m s_k[n-m], \tag{1}$$

where a_m are the linear prediction coefficients. Note that Eq. (1) can also be written as

$$S_k(z) = \frac{1}{1-P(z)} E_k(z)$$
$$P(z) = \sum_{m=1}^{p} a_m z^{-m}$$

Suppose that we wish to exhaustively test K different candidate excitations, $e_k[n]$, for $1 \le k \le K$. We want to choose the excitation that minimizes the perceptually weighted error, $\mathcal{E}_k$, defined as

$$\mathcal{E}_k = \sum_{n=0}^{N-1} y_k^2[n],$$

where

$$Y_k(z) = \frac{1-P(z)}{1-P(z/\alpha)} S_k(z),$$

Demonstrate that $y_k[n]$ can be generated from $e_k[n]$ using only p multiplications per sample.

Solution:

$$Y_k(z) = \frac{1}{1-P(z/\alpha)} E_k(z)$$
$$P(z/\alpha) = \sum_{m=1}^{p} a_m \alpha^m z^{-m}$$

Therefore, if we compute the coefficients $c_m = \alpha^m a_m$ once per frame, we can then compute all of the frame's N samples using

$$y_k[n] = e_k[n] + \sum_{m=1}^{p} c_m y_k[n-m]$$

4. (20 points) The scaling constant, in the standard scaled-forward algorithm, can be interpreted as

$$c_t = P(o_t | o_1, \ldots, o_{t-1}, \lambda)$$

This is an intriguing quantity; it suggests that we are predicting the next spectrum, given the previous spectra. Suppose that somebody else has provided you with a table of the non-scaled forward probabilities for a particular waveform,

$$\alpha_t(i) = P(q_t = i, o_1, \ldots, o_t | \lambda)$$

Is it possible to compute c_T for the last frame without computing the scaled forward algorithm for all time steps? In other words, can you come up with a formula for c_T in terms of $\alpha_t(i)$, $a_{i,j}$, and $b_i(k)$, for some appropriate values of i, j, t, k, but without computing the scaled forward algorithm for all time steps?

Solution: First, we want a probability conditioned on $o_1, \ldots, o_{t-1}$. We can get that by normalizing $\alpha_{t-1}(i)$:

$$P(q_{t-1} = i | o_1, \ldots, o_{t-1}, \lambda) = \frac{\alpha_{t-1}(i)}{\sum_{j=1}^{N} \alpha_{t-1}(j)}$$

Then we can find the probability of o_t given $o_1, \ldots, o_{t-1}$ by summing over all of the ways in which o_t could have been made:

$$\begin{aligned} P(o_t | o_1, \ldots, o_{t-1}, \lambda) &= \sum_{i=1}^{N} \sum_{j=1}^{N} P(q_{t-1} = i | o_1, \ldots, o_{t-1}, \lambda) a_{ij} b_j(o_t) \\ &= \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_{t-1}(i) a_{ij} b_j(o_t)}{\sum_{j=1}^{N} \alpha_{t-1}(j)} \\ &= \frac{\sum_{j=1}^{N} \alpha_t(j)}{\sum_{j=1}^{N} \alpha_{t-1}(j)} \end{aligned}$$

5. (20 points) Recall that Baum's auxiliary can be written as

$$Q(\lambda, \bar{\lambda}) = \sum_{Q} P(Q|O, \lambda) \ln P(O, Q|\bar{\lambda}),$$

and that the part related to the observation pdf can be simplified to

$$Q_b(\lambda, \bar{\lambda}) = \sum_{t=1}^{T} \sum_{i=1}^{N} \gamma_t(i) \ln \bar{b}_i(o_t),$$

where the terms are defined as

$$\gamma_t(i) = P(q_t = i|O, \lambda)$$
$$\bar{b}_i(o_t) = P(o_t = o_t|q_t = i, \bar{\lambda})$$

Suppose that we have a sequence of non-negative scalar observations, $O = [o_1, \ldots, o_T]$, modeled by exponential probability density functions:

$$\bar{b}_i(o_t) = \begin{cases} \frac{1}{\bar{\mu}_i} \exp\left(-o_t/\bar{\mu}_i\right) & o_t \geq 0 \\ 0 & o_t < 0 \end{cases}$$

where $\bar{\mu}_i$ is the state-dependent mean. The exponential pdf is only well defined if $\mu_i \geq 0$. We can force $\bar{\mu}_i$ to be non-negative by maximing a Lagrangian term of the form

$$\mathcal{L}(\bar{\lambda}) = Q_b(\lambda, \bar{\lambda}) - \sum_{i=1}^{N} \kappa_i \bar{\mu}_i$$

Simplify $\mathcal{L}(\bar{\lambda})$ so that it is a function of only $\gamma_t(i)$, o_t, $\bar{\mu}_i$, and κ_i for $1 \leq t \leq T$, $1 \leq i \leq N$.

Solution: This looks complicated, but after a bit of staring, we realize that it is only asking us to take the logarithm of an exponential pdf. The result is

$$\mathcal{L}(\bar{\lambda}) = \sum_{t=1}^{T} \sum_{i=1}^{N} \gamma_t(i) \left(-\ln \bar{\mu}_i - \frac{o_t}{\bar{\mu}_i} \right) - \sum_{i=1}^{N} \kappa_i \bar{\mu}_i$$